

500.42992X00

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

Applicant(s): KAWAMURA, et al.
Serial No.: Not assigned
Filed: July 29, 2003
Title: DATABASE PROCESSING METHOD AND SYSTEM USING LOG
INFORMATION, PROCESSING PROGRAM THEREOF, AND
STORAGE UNIT FOR EXECUTION THEREOF
Group: Not assigned

LETTER CLAIMING RIGHT OF PRIORITY

Mail Stop Patent Application
Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

July 29, 2003

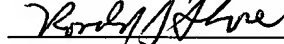
Sir:

Under the provisions of 35 USC 119 and 37 CFR 1.55, the applicant(s) hereby claim(s) the right of priority based on Japanese Application No.(s) 2002-368688 filed December 19, 2002.

A certified copy of said Japanese Application is attached.

Respectfully submitted,

ANTONELLI, TERRY, STOUT & KRAUS, LLP



Ronald J. Shore

Registration No. 28,577

RJS/amr
Attachment
(703) 312-6600

日本国特許庁
JAPAN PATENT OFFICE

別紙添付の書類に記載されている事項は下記の出願書類に記載されている事項と同一であることを証明する。

This is to certify that the annexed is a true copy of the following application as filed with this Office.

出願年月日
Date of Application: 2002年12月19日

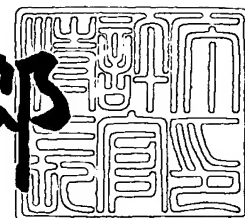
出願番号
Application Number: 特願2002-368688
[ST. 10/C]: [JP 2002-368688]

出願人
Applicant(s): 株式会社日立製作所

2003年 7月 9日

特許庁長官
Commissioner,
Japan Patent Office

太田信一郎



出証番号 出証特2003-3054366

【書類名】 特許願

【整理番号】 K02016651

【あて先】 特許庁長官殿

【国際特許分類】 G06F 12/00

【発明者】

 【住所又は居所】 神奈川県横浜市戸塚区戸塚町 5 0 3 0 番地 株式会社日立製作所 ソフトウェア事業部内

 【氏名】 河村 信男

【発明者】

 【住所又は居所】 千葉県松戸市二十世紀が丘丸山町 1 7

 【氏名】 喜連川 優

【発明者】

 【住所又は居所】 神奈川県小田原市中里 3 2 2 番地 2 号 株式会社日立製作所 S A N ソリューション事業部内

 【氏名】 正井 一夫

【特許出願人】

 【識別番号】 000005108

 【氏名又は名称】 株式会社日立製作所

【代理人】

 【識別番号】 100083552

 【弁理士】

 【氏名又は名称】 秋田 収喜

 【電話番号】 03-3893-6221

【手数料の表示】

 【予納台帳番号】 014579

 【納付金額】 21,000円

【提出物件の目録】

 【物件名】 明細書 1

 【物件名】 図面 1

【物件名】

要約書 1

【プルーフの要否】

要

【書類名】 明細書

【発明の名称】 データベース処理方法及び装置並びにその処理プログラム

【特許請求の範囲】

【請求項 1】 ホストコンピュータのバッファ上で行われたデータベース処理の内容を記憶装置上のデータベースに反映させる処理を行うデータベース処理方法において、

ホストコンピュータから送信されたアクセス要求を受信し、その受信したアクセス要求が書き込み要求または読み込み要求のいずれであるかを判定するステップと、前記受信したアクセス要求が書き込み要求である場合に、その書き込み内容がホストコンピュータのバッファ上で行われたデータベース処理の内容を示すログ情報であるかどうかを判定するステップと、前記書き込み内容が前記ログ情報である場合に、ホストコンピュータ側のデータベース処理で認識している論理的な位置情報と記憶装置サブシステム上の物理的な位置情報との対応関係を示す変換テーブルによって、前記ログ情報中に示された位置情報を記憶装置サブシステム上の物理的な位置情報に変換するステップと、その変換した物理的な位置情報で表される記憶装置サブシステム上のデータベース領域のデータを前記ログ情報の内容に従って更新するステップとを有することを特徴とするデータベース処理方法。

【請求項 2】 ホストコンピュータのバッファ上で行われたデータベース処理の内容を記憶装置上のデータベースに反映させる処理を行うデータベース処理方法において、

ホストコンピュータのデータベースバッファの内容を記憶装置サブシステム内の記憶装置へ反映させる必要が生じた場合に、そのデータベースバッファに対して行われたデータベース処理の内容を示すログ情報の書き込み要求をホストコンピュータから記憶装置サブシステムへ送信するステップと、データベースバッファ中にデータベース処理でアクセス対象となっているデータが存在していない場合に当該データの読み込み要求をホストコンピュータから記憶装置サブシステムへ送信するステップと、

ホストコンピュータから送信されたアクセス要求を受信し、その受信したアク

セス要求が書き込み要求または読み込み要求のいずれであるかを判定するステップと、前記受信したアクセス要求が書き込み要求である場合に、その書き込み内容が前記ログ情報であるかどうかを判定するステップと、前記書き込み内容が前記ログ情報である場合に、ホストコンピュータ側のデータベース処理で認識している論理的な位置情報と記憶装置サブシステム上の物理的な位置情報との対応関係を示す変換テーブルによって、前記ログ情報中に示された位置情報を記憶装置サブシステム上の物理的な位置情報に変換するステップと、その変換した物理的な位置情報で表される記憶装置サブシステム上のデータベース領域のデータを前記ログ情報の内容に従って更新するステップとを有することを特徴とするデータベース処理方法。

【請求項 3】 前記受信したアクセス要求が読み込み要求である場合に、それ以前の書き込み要求で受信したログ情報中に読み込み対象のデータを更新するログ情報が含まれているかどうかを判定し、前記受信したログ情報中に読み込み対象のデータを更新するログ情報が含まれている場合に、読み込み対象のデータを当該ログ情報の内容に従って更新することを特徴とする請求項 1 または請求項 2 のいずれかに記載されたデータベース処理方法。

【請求項 4】 前記ログ情報の内、COMMIT されたトランザクションのログ情報を用いて更新を行うことを特徴とする請求項 1 乃至請求項 3 のいずれか 1 項に記載されたデータベース処理方法。

【請求項 5】 前記データベース領域のデータの更新を、そのデータベース領域のデータに対応する物理デバイス毎に並列に行うことを特徴とする請求項 1 乃至請求項 4 のいずれか 1 項に記載されたデータベース処理方法。

【請求項 6】 前記ログ情報の内容に従って更新されるデータベース領域のデータが他の記憶装置サブシステム中に存在している場合に、そのログ情報の内容を当該記憶装置サブシステムへ送信し、その記憶装置サブシステム内でデータベース領域のデータの更新を行うことを特徴とする請求項 1 乃至請求項 5 のいずれか 1 項に記載されたデータベース処理方法。

【請求項 7】 ホストコンピュータのバッファ上で行われたデータベース処理の内容を記憶装置上のデータベースに反映させる処理を行う記憶装置サブシ

テムにおいて、

ホストコンピュータから送信されたアクセス要求を受信し、その受信したアクセス要求が書き込み要求または読み込み要求のいずれであるかを判定する制御処理部と、

前記受信したアクセス要求が書き込み要求である場合に、その書き込み内容がホストコンピュータのバッファ上で行われたデータベース処理の内容を示すログ情報であるかどうかを判定し、前記書き込み内容が前記ログ情報である場合に、ホストコンピュータ側のデータベース処理で認識している論理的な位置情報と記憶装置サブシステム上の物理的な位置情報との対応関係を示す変換テーブルによって、前記ログ情報中に示された位置情報を記憶装置サブシステム上の物理的な位置情報に変換し、その変換した物理的な位置情報で表される記憶装置サブシステム上のデータベース領域のデータを前記ログ情報の内容に従って更新する更新処理部とを備えることを特徴とする記憶装置サブシステム。

【請求項 8】 前記更新処理部は、受信したアクセス要求が読み込み要求である場合に、それ以前の書き込み要求で受信したログ情報中に読み込み対象のデータを更新するログ情報が含まれているかどうかを判定し、前記受信したログ情報中に読み込み対象のデータを更新するログ情報が含まれている場合に、読み込み対象のデータを当該ログ情報の内容に従って更新するものであることを特徴とする請求項 7 に記載された記憶装置サブシステム。

【請求項 9】 ホストコンピュータのバッファ上で行われたデータベース処理の内容を記憶装置上のデータベースに反映させる処理を行うデータベース処理システムにおいて、

ホストコンピュータのデータベースバッファの内容を記憶装置サブシステム内の記憶装置へ反映させる必要が生じた場合に、そのデータベースバッファに対して行われたデータベース処理の内容を示すログ情報の書き込み要求をホストコンピュータから記憶装置サブシステムへ送信し、データベースバッファ中にデータベース処理でアクセス対象となっているデータが存在していない場合に当該データの読み込み要求をホストコンピュータから記憶装置サブシステムへ送信するデータベース管理処理部と、

ホストコンピュータから送信されたアクセス要求を受信し、その受信したアクセス要求が書き込み要求または読み込み要求のいずれであるかを判定する制御処理部と、

前記受信したアクセス要求が書き込み要求である場合に、その書き込み内容が前記ログ情報であるかどうかを判定し、前記書き込み内容が前記ログ情報である場合に、ホストコンピュータ側のデータベース処理で認識している論理的な位置情報と記憶装置サブシステム上の物理的な位置情報との対応関係を示す変換テーブルによって、前記ログ情報中に示された位置情報を記憶装置サブシステム上の物理的な位置情報に変換し、その変換した物理的な位置情報で表される記憶装置サブシステム上のデータベース領域のデータを前記ログ情報の内容に従って更新する更新処理部とを備えることを特徴とするデータベース処理システム。

【請求項 10】 ホストコンピュータのバッファ上で行われたデータベース処理の内容を記憶装置上のデータベースに反映させる処理を行う記憶装置サブシステムとしてコンピュータを機能させる為のプログラムにおいて、

ホストコンピュータから送信されたアクセス要求を受信し、その受信したアクセス要求が書き込み要求または読み込み要求のいずれであるかを判定する制御処理部と、

前記受信したアクセス要求が書き込み要求である場合に、その書き込み内容がホストコンピュータのバッファ上で行われたデータベース処理の内容を示すログ情報であるかどうかを判定し、前記書き込み内容が前記ログ情報である場合に、ホストコンピュータ側のデータベース処理で認識している論理的な位置情報と記憶装置サブシステム上の物理的な位置情報との対応関係を示す変換テーブルによって、前記ログ情報中に示された位置情報を記憶装置サブシステム上の物理的な位置情報に変換し、その変換した物理的な位置情報で表される記憶装置サブシステム上のデータベース領域のデータを前記ログ情報の内容に従って更新する更新処理部としてコンピュータを機能させることを特徴とするプログラム。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】

本発明はホストコンピュータと記憶装置サブシステムとを接続してデータベース処理を行うデータベース処理システムに関し、特にホストコンピュータのデータベース処理の内容を示すログ情報により、記憶装置サブシステム内のデータベースにデータベース処理内容を反映するデータベース処理システムに適用して有効な技術に関するものである。

【0002】

【従来の技術】

従来のデータベース管理システムでは、ホストコンピュータ上と記憶装置サブシステムとの間でデータベースブロック及びログブロックの入出力を行っている。すなわち、データベース管理システムが入出力の効率を向上させる為にホストコンピュータのメインメモリ上にデータベースのバッファを設定し、当該バッファ上に記憶装置（メインメモリと比較して低速で大容量の磁気ディスク装置等の記憶装置を指すものとする）から入力したデータベースブロックをキャッシングすることによって、できるだけ記憶装置からの入出力を削減している。

【0003】

この様なデータベース管理システムでは、データの更新処理については、予めデータベースバッファ上に入力したデータベースブロックに対して行い、その際の更新履歴情報をログ専用のバッファにログとして書き出した後、トランザクションの決着時にそのログを記憶装置に強制出力することによって整合性を保証している。その際、データベースブロックの記憶装置への反映（書き込み）では、当該ブロックに対して行った更新履歴ログを先行して記憶装置に出力する、所謂 WAL (Write Ahead Log) を遵守する必要がある。

【0004】

また、データベース管理システムでは、障害に備え、定期的にDBの整合性を保証する為、稼動中にチェックポイントを取得する。チェックポイントは、システム障害時の再開始処理を行う起点となるデータベースの整合性が保証されたポイントとなる。主にチェックポイントは稼動中に出力したログブロック数がある一定の回数に達した時点に取得する場合が多い。チェックポイント処理では、その時点のデータベースバッファ上の更新が行われたデータベースブロックを記憶

装置上に全て書き出す処理が行われる（例えば非特許文献 1 参照）。

【0005】

【非特許文献 1】

ジム・グレイ (Jim Gray)、アンドレアス・ロイター (Andreas Reuter) 著、「トランザクション・プロセッシング：コンセプト・アンド・テクニック (TRANSACTION PROCESSING: CONCEPTS AND TECHNIQUES)」(米国)、第1版、モーガン・カウフマン・パブリッシャー (MORGAN KAUFMAN PUBLISHERS) 社発行、1993年、p. 556-557、pp. 604-609

【0006】

【発明が解決しようとする課題】

従来のデータベース管理システムにおけるデータベースブロックの記憶装置への書き込み方法では、前記の様にデータベースバッファ上の更新が行われたデータベースブロックを記憶装置上に全て書き出す処理が行われるが、更新が行われたデータベースブロック中には更新の行われていないレコードも含まれている為、必要の無い入出力も発生し、ホストコンピュータと記憶装置サブシステムとの間の入出力処理に大きな負荷がかかるという問題がある。

【0007】

本発明の目的は上記問題を解決し、ホストコンピュータのバッファ上で行われたデータベース処理の内容を記憶装置サブシステム上のデータベース領域へ反映させる際に、ホストコンピュータと記憶装置サブシステムとの間の入出力処理負荷を低減させることが可能な技術を提供することにある。

【0008】

本発明の他の目的は、ホストコンピュータの障害回復処理時に記憶装置サブシステムのキャッシュメモリ上に上記データベースの反映データが格納されているため、上記回復処理を高速に実行することができる技術を提供することにある。

【0009】

本発明の他の目的はホストコンピュータのバッファ上で行われたデータベース処理の内容をログ情報によって記憶装置サブシステム上のデータベース領域へ反映させる際に、不要な入出力処理を省略することが可能な技術を提供することにある。

ある。

【0010】

本発明の他の目的はホストコンピュータのバッファ上で行われたデータベース処理の内容をログ情報によって記憶装置サブシステム上のデータベース領域へ反映させる際に、その処理を効率的に行うことが可能な技術を提供することにある。

【0011】

本発明の他の目的は複数台の記憶装置サブシステムにデータベース領域を配置している大規模なデータベース処理システムで、ホストコンピュータのバッファ上で行われたデータベース処理の内容を記憶装置サブシステム上のデータベース領域へ反映させる際にも、ホストコンピュータと記憶装置サブシステムとの間の入出力処理負荷を低減させることが可能な技術を提供することにある。

【0012】

【課題を解決するための手段】

本発明は、ホストコンピュータのバッファ上で行われたデータベース処理の内容を記憶装置上のデータベースに反映させる処理を行うデータベース処理システムにおいて、ホストコンピュータで行われたデータベース処理の内容を示すログ情報に従って記憶装置サブシステム上のデータベース領域のデータへデータベース処理内容を反映するものである。

【0013】

本発明のデータベース処理システムでは、記憶装置サブシステム内のデータベース領域の内容を一時的に保持するデータベースバッファと、データベースバッファに対する更新処理の内容を一時的に保持するログバッファとをホストコンピュータに備えており、ホストコンピュータでのデータベース処理の実行に伴ってデータベースバッファの内容が変更され、その変更内容を記憶装置サブシステム内のデータベース領域に反映させる必要が生じた場合には、データベースバッファ上で行われた更新処理の内容を示すログ情報の記憶装置サブシステムへの書き込みを要求するアクセス要求をホストコンピュータから記憶装置サブシステムへ送信する。

【0014】

記憶装置サブシステムでは、前記アクセス要求をホストコンピュータから受信し、その受信したアクセス要求が書き込み要求または読み込み要求のいずれであるかを判定する。そして、その受信したアクセス要求が書き込み要求である場合には、その書き込み内容がホストコンピュータで行われたデータベース処理の内容を示すログ情報であるかどうかを判定する。

【0015】

前記判定の結果、その書き込み内容が前記ログ情報である場合には、ホストコンピュータ側のデータベース処理で認識している論理的な位置情報と記憶装置サブシステム上の物理的な位置情報との対応関係を示す変換テーブルを参照し、前記ログ情報中に示された位置情報を記憶装置サブシステム上の物理的な位置情報に変換した後、その変換した物理的な位置情報で表される記憶装置サブシステム上のデータベース領域のデータを前記ログ情報の内容に従って更新する。

【0016】

そのときに、記憶装置サブシステムのキャッシュメモリに当該反映対象データが格納されている場合には、キャッシュメモリに格納されている当該反映対象データを反映する。キャッシュメモリに格納されていない場合は、キャッシュメモリに反映対象データをアップロードし、その後、上記反映対象データをキャッシュメモリに格納する。

【0017】

またホストコンピュータは、データベースバッファ中にデータベース処理でアクセス対象となっているデータが存在していない場合に、当該データの読み込み要求をホストコンピュータから記憶装置サブシステムへ送信し、記憶装置サブシステムは、ホストコンピュータから受信したアクセス要求が読み込み要求である場合には、それ以前の書き込み要求で受信したログ情報中に、読み込み対象のデータを更新するログ情報が含まれているかどうかを判定し、そのログ情報中に読み込み対象のデータを更新するログ情報が含まれている場合には、読み込み対象のデータを当該ログ情報の内容に従って更新してホストコンピュータ側へ送信する。

【0018】

前記の様に本発明では、データベースバッファの内容を記憶装置サブシステム内のデータベース領域に反映させる必要が生じた場合に、更新処理の行われたレコードを1つでも含むデータベースブロック全てをホストコンピュータから記憶装置サブシステムへ送信するのではなく、データベースバッファに対する更新処理の内容を示すログ情報をホストコンピュータから記憶装置サブシステムへ送信するので、ホストコンピュータと記憶装置サブシステムとの間で送信されるデータ量を減少させることができる。

【0019】

以上の様に本発明のデータベース処理システムによれば、ホストコンピュータのバッファ上で行われたデータベース処理の内容を示すログ情報に従って記憶装置サブシステム上のデータベース領域のデータを更新するので、ホストコンピュータのバッファ上で行われたデータベース処理の内容を記憶装置サブシステム上のデータベース領域へ反映させる際に、ホストコンピュータと記憶装置サブシステムとの間の入出力処理負荷を低減させることが可能である。

【0020】**【発明の実施の形態】****(実施形態1)**

以下にホストコンピュータのバッファ上で行われたデータベース処理の内容をディスクサブシステムの磁気ディスク装置上のデータベースに反映させる処理を行う実施形態1のデータベース処理システムについて説明する。

【0021】

図1は本実施形態のデータベース処理システムのシステム構成を示す図である。図1に示す様に本実施形態のホストコンピュータ1はデータベース管理処理部10を有している。データベース管理処理部10は、ホストコンピュータ1のDBバッファ12の内容をディスクサブシステム2内の磁気ディスク装置へ反映させる必要が生じた場合に、そのDBバッファ12に対して行われたデータベース処理の内容を示すログ情報であるログブロック262aの書き込み要求をホストコンピュータ1からディスクサブシステム2へ送信し、DBバッファ12中にデ

データベース処理でアクセス対象となっているデータが存在していない場合に当該データの読み込み要求をホストコンピュータ 1 からディスクサブシステム 2 へ送信する処理部である。

【0022】

ホストコンピュータ 1 をデータベース管理処理部 10 として機能させる為のプログラムは、CD-ROM等の記録媒体に記録され磁気ディスク等に格納された後、メモリにロードされて実行されるものとする。なお前記プログラムを記録する記録媒体はCD-ROM以外の他の記録媒体でも良い。また前記プログラムを当該記録媒体から情報処理装置にインストールして使用しても良いし、ネットワークを通じて当該記録媒体にアクセスして前記プログラムを使用するものとしても良い。

またディスクサブシステム 2 は、ディスク制御処理部 21 と、ディスクアクセス制御部 23 と、更新処理部 30 とを有している。

【0023】

ディスク制御処理部 21 は、ホストコンピュータ 1 から送信されたアクセス要求を受信し、その受信したアクセス要求が書き込み要求または読み込み要求のいずれであるかを判定する処理や、ディスクサブシステム装置全体の動作を制御する制御処理部である。

【0024】

ディスクアクセス制御部 23 は、ディスクサブシステム 2 配下の各磁気ディスク装置へのアクセスを制御する処理部である。更新処理部 30 は、前記受信したアクセス要求が書き込み要求である場合に、その書き込み内容がホストコンピュータ 1 のDBバッファ 12 上で行われたデータベース処理の内容を示すログ情報であるかどうかを判定し、前記書き込み内容が前記ログ情報である場合に、ホストコンピュータ 1 側のデータベース処理で認識している論理的な位置情報とディスクサブシステム 2 上の物理的な位置情報との対応関係を示すDB-ディスクブロック変換テーブル 28 によって、前記ログ情報中に示された位置情報をディスクサブシステム 2 上の物理的な位置情報に変換し、その変換した物理的な位置情報で表されるディスクサブシステム 2 上のデータベース領域 24 のデータを前記

ログ情報の内容に従って更新する処理部である。

【0025】

ディスクサブシステム 2 を、ディスク制御処理部 21、ディスクアクセス制御部 23 及び更新処理部 30 として機能させる為のプログラムは、フロッピーディスク等の記録媒体に記録されて実行されるものとする。なお前記プログラムを記録する記録媒体はフロッピーディスク以外の他の記録媒体でも良い。また前記プログラムを当該記録媒体から情報処理装置にインストールして使用しても良いし、ネットワークを通じて当該記録媒体にアクセスして前記プログラムを使用するものとしても良い。

【0026】

本実施形態のホストコンピュータ 1 とディスクサブシステム 2 は、ストレージエリアネットワーク 16 で接続されている。

ホストコンピュータ 1 では、データベース管理処理部 10 が稼動し、ホストコンピュータ 1 は、ディスクサブシステム 2 内のデータベース領域 24 の内容を一時的に保持する DB バッファ 12 と、DB バッファ 12 に対する更新処理の内容を一時的に保持するログバッファ 14 とを備えている。

【0027】

ディスクサブシステム 2 では、ホストコンピュータ 1 からの命令を受けて動作するディスク制御処理部 21 と、キャッシュメモリ 22 と、ディスクアクセス制御部 23 とを通して磁気ディスク装置上のデータベース領域 24 へのアクセスが行われており、ディスクアクセスは常にキャッシュメモリ 22 を介して行われることになる。

【0028】

本実施形態では、ホストコンピュータ 1 上のデータベース管理処理部 10 で動作するトランザクションによって更新されたデータの更新履歴情報であるログ情報がログブロック 262a に書き込まれ、トランザクションの決着時等、DB バッファ 12 の内容をディスクサブシステム 2 へ反映させる必要が生じた場合に、ログブロック 262a がディスクサブシステム 2 に書き込まれる。

【0029】

本実施形態の更新処理部 30 は、データの書き込みが行われると、そのデータがログブロック 262a であるかどうかを判定し、ディスクサブシステム 2 上でデータベースブロック 242a の書き込みを制御する。

【0030】

すなわち更新処理部 30 は、ディスクサブシステム 2 が受け付けたコマンドを解析し、ログブロック 262a の書き込み要求であれば、キャッシュメモリ 22 上に書き込んだログブロック 262a を解析し、ログブロック 262a 中のログレコードから該当するデータベース領域 24 のデータベースブロック 242a をキャッシュメモリ 22 上にアップロードし、ログの内容を反映する処理を行う。

【0031】

このとき反映されるデータベースブロック 242a は、ログブロック 262a 中のログレコードではホストコンピュータ 1 上のデータベース管理処理部 10 が認識できる論理的な位置情報で表されている為、この論理的な位置情報をディスクサブシステム 2 上の物理的な位置情報にマッピングする必要がある。そこで、この処理を DB-ディスクブロック変換テーブル 28 を用いて行う。DB-ディスクブロック変換テーブル 28 は、一般的には DB 構築時にホストコンピュータ 1 上のデータベース管理処理部 10 から作成されることになる。

【0032】

図 2 は本実施形態の DB-ディスクブロック変換テーブル 28 の構成情報を示す図である。図 2 に示す様に DB-ディスクブロック変換テーブル 28 は、データベース領域 24 を識別する為の情報であるデータベース領域 ID と、そのデータベース領域 ID で識別されるデータベース領域が複数のファイルで構成される場合のファイルの順序番号を示すファイル ID と、前記データベース領域を構成するブロックの長さを示すブロック長と、前記データベース領域の構成ファイルが確保されている論理ボリュームを識別する為の情報である論理ボリューム ID と、その論理ボリューム ID で識別される論理ボリュームがマッピングされているディスクサブシステムを識別する為の番号であるディスク制御装置番号と、そのディスク制御装置番号で識別されるディスクサブシステムの磁気ディスク装置の中で、前記論理ボリュームがマッピングされている磁気ディスク装置のドライ

ブ番号を識別する為の情報である物理デバイスIDと、その物理デバイスIDで識別される磁気ディスク装置上での当該ファイルの相対的な位置を示す相対位置とを格納している。

【0033】

データベース領域24を構成するファイルは、ホストコンピュータ1上のオペレーティングシステムが認識するファイルシステムとして論理ボリュームにマッピングされる。更に、論理ボリュームは、ディスクサブシステム2の物理デバイスである磁気ディスク装置に対応したデバイスファイルとしてマッピングされる。

【0034】

ディスクサブシステム2内では、前記デバイスファイルは、LU (Logical Unit) に対応している。従って、データベース領域24を構成するファイルは、最終的に物理デバイスである磁気ディスク装置にマッピングされる。対応する物理情報は、ディスクサブシステム2上の物理デバイスを識別する為の物理デバイスIDと、物理デバイス内の相対位置であるLBA (Logical Block Address) である。

【0035】

図3は本実施形態のホストコンピュータ上で認識されるデータベース領域と、オペレーティングシステムが認識する論理ボリュームと、デバイスファイル及びディスクサブシステム内のLUへのマッピング関連の例を示す図である。図3に示す様にデータベース管理処理部10では、データを格納するデータベース領域は、複数のファイルから構成されるものとして認識されている。構成する各ファイルは、ホストコンピュータ1上のオペレーティングシステムのファイルに対応しており、図3ではオペレーティングシステムにおいてRAWデバイスとして認識されるケースを想定している。更に、オペレーティングシステムのファイルは、物理的な磁気ディスク装置に対応するデバイスファイルとして管理されており、そのデバイスファイルは、前述した様にディスクサブシステム2内の各々の磁気ディスク装置に対応するLUにマッピングされている。

【0036】

次に図4から図7に示す流れ図を用いてホストコンピュータ1上のデータベース管理処理部10からログバッファ14中のログブロック262aのディスクサブシステム2への書き込みを要求するアクセス要求の処理について説明する。始めに、図4を用いて処理の概要を示す。

【0037】

図4は本実施形態の受信コマンド解析処理の処理手順を示すフローチャートである。図4に示した処理は、ディスクサブシステム2内のプロセッサによって実行されるディスク制御処理部21の処理として実現されるものであり、ホストコンピュータ1からのアクセス要求を受領したディスクサブシステム2は、始めに受領コマンドの解析処理を行う（ステップ300）。接続チャネルのプロトコルに従ってコマンドを解析することで、アクセス要求がREADコマンドであるかWRITEコマンドであるかを識別できるものとする。

【0038】

ステップ320でディスク制御処理部21は、受領コマンドがWRITEコマンドであるかを判定し、WRITEコマンドである場合にはWRITEコマンド処理を行う（ステップ340）。また、READコマンドである場合にはREADコマンド処理（ステップ360）を行う。

【0039】

図5は本実施形態のWRITEコマンド受領時の処理手順を示すフローチャートである。図5の様に更新処理部30は、ディスク制御処理部21からコマンドを受領すると、受信コマンドからコマンド種類とアクセス先アドレスを解析し、WRITEコマンドであることを認識する（ステップ341）。ここで、アクセス先アドレスからは、複数のディスクサブシステムやその各磁気ディスク装置に割り当てられているアドレスを示す装置構成管理テーブルの情報との比較を行うことにより、アクセス要求先のディスク制御装置番号とドライブ番号を識別することができるものとする。

【0040】

次に、ステップ341で解析したアクセス先アドレスのデータが、ディスクサブシステム2のキャッシュメモリ22に保持されているかどうかを調べ、キャッ

シュヒットミス判定を行う（ステップ342）。

【0041】

アクセス先データがキャッシュメモリ22に保持されていないキャッシュミスの場合には、前記の様にアクセス要求先のドライブ番号を識別し、ディスクサブシステム2のディスクアクセス制御部23に対してそのドライブ番号に対応する磁気ディスク装置からキャッシュメモリ22への転送依頼を行う（ステップ343）。この場合、転送終了までWRITE処理を中断し（ステップ344）、転送処理終了後、再度WRITE処理を継続する。また、転送先のキャッシュアドレスはキャッシュの空きリスト等、一般的な方法で管理、取得すれば良いが、転送先アドレスについては、キャッシュ管理テーブルを更新することで登録する必要がある。

【0042】

ステップ342でキャッシュヒットの判定の場合、またはステップ344で転送処理が終了した場合には、ディスクサブシステム2内のキャッシュメモリ22に対して当該データの更新を行う（ステップ345）。すなわち、ホストコンピュータ1から受領したデータの内容を書き込む。

【0043】

データの更新が終了した後、前記アクセス先アドレスがログ用ディスク26内のアドレスであるかどうかを調べて当該データがデータベース処理のログ用ディスク26のデータであるかを判定し（ステップ346）、書き込み内容がログ用ディスク26へのデータ、すなわちログブロックである場合には、そのログブロックの内容に従って該当するデータベース領域24のデータベースブロックへのログ追跡処理を行う（ステップ347）。

【0044】

ログ追跡処理が完了するか、ステップ346の判定でログブロックではないと判定された場合、ホストコンピュータ1に対してWRITEコマンド処理の完了報告を行う（ステップ348）。

【0045】

図6は本実施形態のログ追跡処理の処理手順を示すフローチャートである。ロ

グブロックは複数のログレコードの集まりである。従って、図6に示す様にログブロックに含まれるログレコードについて順次、処理を行っていく。

【0046】

まず更新処理部30は、ログレコードのログ情報がトランザクションの開始処理、COMMITまたはROLLBACKといった決着処理を示す情報であるかを判定する（ステップ401）。

【0047】

当該ログレコードがトランザクションの状態変更ログではない場合には、更にデータベースの更新履歴を示すトランザクション更新ログであるかを判定する（ステップ402）。

【0048】

当該ログレコードがトランザクション更新ログである場合には、図2に示したDB-ディスクブロック変換テーブル28を参照し、ログレコード中に記録されたログ情報に含まれているデータベース領域ID、ファイルID、ページ番号から、対応する物理ディスクのディスク制御装置番号とドライブ番号とページ番号を識別する（ステップ403）。すなわち、当該ログ情報に含まれているデータベース領域ID及びファイルIDに一致するレコードを、DB-ディスクブロック変換テーブル28から検索して該当するディスク制御装置番号、ドライブ番号及び相対位置を求めた後、その相対位置がファイルの先頭であるものとして、当該ログ情報のページ番号を物理ディスク上のページ番号に変換する。

【0049】

次にステップ403で前記の様に識別したデータがキャッシュメモリ22に保持されているかどうかを調べてキャッシュヒットミス判定を行う（ステップ404）。当該データがキャッシュメモリ22に保持されていないキャッシュミスの場合には、ディスクアクセス制御部23に対して当該データベースブロックのドライブからキャッシュメモリ22への転送依頼を行う（ステップ405）。

【0050】

ステップ404で当該データベースブロックがキャッシュヒットするか、ステップ405で転送処理が終了するとキャッシュメモリ22中の当該データベース

ブロックに対してログレコードに含まれているデータベース更新履歴情報を反映する（ステップ406）。

【0051】

一方、ステップ401で判定した結果、当該ログレコードがトランザクションの状態変更ログであり、ROLLBACKログである場合には、ステップ408で当該トランザクションによる更新履歴情報の反映を取り消す処理を行う。

【0052】

ステップ407では、当該ログブロックの全ログレコードの処理を完了したかどうかを調べ、まだ全ログレコードの処理を完了していない場合にはステップ401へ戻り、完了した場合には処理を終了する。

【0053】

図7は本実施形態のREADコマンド受領時の処理手順を示すフローチャートである。図7の様に更新処理部30は、ディスク制御処理部21からコマンドを受領すると、受信コマンドからコマンドの種類とアクセス先アドレスを解析し、READアクセス要求であることを認識する（ステップ361）。ここで、アクセス先アドレスからは、前記と同様にして装置構成管理テーブルを参照することで、アクセス要求先のディスク制御装置番号とドライブ番号を識別できるものとする。

【0054】

次に、ステップ361で解析したアクセス先アドレスのデータが、ディスクサブシステム2のキャッシュメモリ22に保持されているかどうかを調べ、キャッシュヒットミス判定を行う（ステップ362）。

【0055】

アクセス先データがキャッシュメモリ22に保持されていないキャッシュミスの場合には、前記の様にアクセス要求先のドライブ番号を識別し、ディスクサブシステム2のディスクアクセス制御部23に対してそのドライブ番号に対応する磁気ディスク装置からキャッシュメモリ22への転送依頼を行う（ステップ363）。この場合、転送終了までREAD処理を中断し（ステップ364）、転送処理終了後、再度READ処理を継続する。また、転送先のキャッシュアドレス

はキャッシュの空きリスト等、一般的な方法で管理、取得すれば良いが、転送先アドレスについては、キャッシュ管理テーブルを更新することで登録する必要がある。

【0056】

ステップ362でキャッシュヒットの判定の場合、またはステップ364で転送処理が終了した場合、従来の単純なデータの読み出しの場合では当該ディスクサブシステム内のキャッシュメモリのデータをチャネルに転送するが、本実施形態では、更に当該データがデータベース管理処理部10からのデータベースブロックのREAD要求であるかを判定する(ステップ365)。当該データがデータベースブロックであるかは、DB-ディスクブロック変換テーブル28を参照し、該当ドライブ番号の存在を判定することによって識別できる。

【0057】

当該データがデータベースブロックである場合には、それ以前のWRITE要求で受信し、ログ追跡処理の終了していないログ情報中にそのデータベースブロックを更新するログレコードが含まれているかどうかを判定し、含まれている場合にはその更新を行う。

【0058】

すなわち、そのREADアクセス要求先の物理ドライブのドライブ番号とページ番号をアクセス先アドレスから求め、ログ追跡処理の終了していないログレコードの物理ドライブのドライブ番号及びページ番号と比較して、キャッシュメモリ22中のログブロック262a内のログレコード中に反映する必要のあるログレコードがあるかを判定し(ステップ366)、該当ログレコードが存在した場合にはログ追跡処理を行う(ステップ367)。

そして、ステップ365の処理によって該当データがデータベースブロックでないと判定された時点か、若しくはステップ367のログ追跡処理が完了した時点で該当データをチャネルに転送する。

【0059】

前記の様に本実施形態では、DBバッファ12に対する更新処理の内容を示すログ情報によりディスクサブシステム2上のデータベース領域24のデータを更

新するので、従来のデータベース管理システムで行っていたデータベースブロックの記憶装置サブシステムへの書き込みについては本実施形態では必要無くなる。また、チェックポイント取得時にも同様にログ情報によりデータベース領域 24 への書き込みをディスクサブシステム 2 側で行う。この為、本実施形態においてホストコンピュータ 1 は、チェックポイント取得処理等の DB バッファ 12 の内容をディスクサブシステム 2 へ反映させる処理を瞬時に終了させることが可能であり、ログ情報によりディスクサブシステム 2 上でデータベース領域 24 のデータを更新している間も、ホストコンピュータ 1 側ではデータベース処理を続行することができる。

【0060】

この際、ホストコンピュータ 1 でデータベース処理を続行中にアクセス対象のデータが DB バッファ 12 中に存在していないことが検出され、ディスクサブシステム 2 に対してデータベースブロックの読み込み要求が行われた場合には、その読み込み対象のデータベースブロックをログ情報の内容に従って更新した後にホストコンピュータ 1 へ送信するので、ホストコンピュータ 1 は、ディスクサブシステム 2 でのログ追跡処理を何ら意識すること無くデータベース処理を続行することが可能である。

【0061】

更に本実施形態では、ホストコンピュータ 1 からディスクサブシステム 2 へのデータベースブロックの書き込みが不要になることによって、ディスクサブシステム 2 に対する帯域幅を大幅に増加させたのと同様の効果を得ることが可能になる。すなわち本実施形態では、DB バッファ 12 の内容をディスクサブシステム 2 内のデータベース領域 24 に反映させる必要が生じた場合に、更新処理の行われたレコードを 1 つでも含むデータベースブロック 242 a 全てをディスクサブシステム 2 へ送信するのではなく、DB バッファ 12 に対する更新処理の内容を示すログブロック 262 a をディスクサブシステム 2 へ送信するので、ディスクサブシステム 2 へ送信されるデータ量を減少させることが可能であり、この為、ディスクサブシステム 2 に対する帯域幅を相対的に増加させることができる。

【0062】

一方、ホストコンピュータ 1 側のデータベース管理処理が障害によってダウンした場合であっても、ディスクサブシステム 2 上のキャッシュメモリ 22 は、最新のデータベースブロック 242 の状態が保持されたウォームキャッシュ状態となっているので、再開始処理時にホストコンピュータ 1 からディスクサブシステム 2 に対する入出力要求があった場合にキャッシュヒットとなり、実際に磁気ディスク装置上のデータベース領域 24 までアクセスする頻度を極端に軽減することができる。

【0063】

以上説明した様に本実施形態のデータベース処理システムによれば、ホストコンピュータのバッファ上で行われたデータベース処理の内容を示すログ情報に従って記憶装置サブシステム上のデータベース領域のデータを更新するので、ホストコンピュータのバッファ上で行われたデータベース処理の内容を記憶装置サブシステム上のデータベース領域へ反映させる際に、ホストコンピュータと記憶装置サブシステムとの間の入出力処理負荷を低減させることが可能である。

【0064】

データベース領域へ反映したデータは、記憶装置サブシステム内のキャッシュに格納されているため、反映データを利用する場合や、特にデータを大量に参照するホストコンピュータの障害回復処理の高速化を図ることが可能である。

【0065】

(実施形態 2)

以下にログ情報の内、COMMITされたトランザクションのログ情報を用いて更新を行う実施形態 2 のデータベース処理システムについて説明する。

実施形態 1 のログ追跡処理では、全てのログレコードを対象に該当するデータベースブロックへの反映処理を行ったが、本実施形態ではログ追跡処理の別の実施方法について図 8 から図 10 を用いて説明する。

【0066】

図 8 は本実施形態のログ追跡処理においてログブロック中に含まれる全てのログレコードについてトランザクション毎にログを解析した結果の例を示す図である。図 8 に示す様に本実施形態では、ログブロック 262a を解析し、まず当該

ディスクサブシステム内のキャッシュメモリ 22 とは異なる共用メモリ上に抽出ログバッファ 264 を確保して、抽出ログバッファ 264 に全てのログレコード 266a～266l を格納する。

【0067】

このとき、各ログレコードをトランザクション別に管理する為、トランザクションログ管理テーブル 268 によって各々のログレコードをトランザクション別に管理し、トランザクション TR1 からトランザクション TR4 までのログレコードのチェーンをそれぞれ生成する。

【0068】

すなわち、トランザクション TR1 には、トランザクションログ管理テーブル 268a、268b、268c、乃至 268f がチェーンされ、トランザクション TR2 には、トランザクションログ管理テーブル 268e、268g がチェーンされる。またトランザクション TR3 には、トランザクションログ管理テーブル 268h、268j、268l がチェーンされ、トランザクション TR4 には、トランザクションログ管理テーブル 268i、268k がチェーンされる。

【0069】

この様に、トランザクションログ管理テーブル 268 を各々のトランザクション毎にチェーンさせることにより、ログレコード情報にトランザクションの正常決着処理を示す COMMIT が識別されたトランザクションのログレコードだけを対象にすることができる。

【0070】

図9は本実施形態の図8のトランザクションログ管理テーブル 268 を用いる際のログレコード分別処理の処理手順を示すフローチャートである。本処理は、図5のステップ 347 で示した WRITE コマンド処理に代わる処理である。

【0071】

ログブロック中の各ログレコードについて、トランザクション開始を示すトランザクション BEGIN ログであるかを判定し（ステップ 441）、トランザクション BEGIN ログである場合、そのログを抽出ログバッファ 264 へ追加し、トランザクションログ管理テーブル 268 への登録を行う。

【0072】

ステップ441の判定でトランザクションBEGINログではないと判定された場合にはデータベース更新ログであるかを判定し（ステップ443）、データベース更新ログである場合には、該当するトランザクション識別子と同じトランザクションログ管理テーブル268の最後尾にチェーンする。

【0073】

ステップ443の判定でトランザクション更新ログではないと判定された場合には、トランザクションの無効化を示すトランザクションROLLBACKログかであるかを判定し（ステップ445）、トランザクションROLLBACK処理である場合には、該当するトランザクション識別子と同じ識別子のトランザクションログ管理テーブル268を削除すると同時に抽出口バッファ264の該当するログレコードも削除する。すなわち、ROLLBACKしたトランザクションのログは、データベースブロックへ反映されない様にする。

【0074】

ステップ445の判定でトランザクションROLLBACKログではないと判定された場合には、トランザクションの有効化を示すトランザクションCOMMITログであるかを判定し（ステップ447）、トランザクションCOMMITログである場合にはログ追跡処理を行う（ステップ448）。

【0075】

ステップ447の判定でトランザクションCOMMITログではないと判定された後、ステップ449でログブロックの終了を検出するまでステップ441からステップ449までを繰り返す。

【0076】

図10は本実施形態の図9のステップ448におけるログ追跡処理の処理手順を示すフローチャートである。図10のログ追跡処理では、COMMITしたトランザクションのトランザクションログ管理テーブル268の先頭アドレスの次のアドレスが引き渡される。つまり、トランザクションBEGINログを処理の対象から外することができる。

【0077】

この処理では、1つのトランザクションの複数のログレコードの集まりについて順次、処理を行っていく。まずステップ4481では、ログレコードのログ情報がトランザクションのCOMMITログであるかどうかを判定し、当該ログレコードが、COMMITログではなく、トランザクション更新ログである場合には、ログレコード中に記録された当該ログ情報に含まれているデータベース領域ID、ファイルID及びページ番号と、図2に示したDB-ディスクブロック変換テーブル28の情報とを比較し、対応する物理ディスクのディスク制御装置番号とドライブ番号とページ番号を識別する（ステップ4482）。

【0078】

次にステップ4481で識別した当該データについてキャッシュメモリ22に対してキャッシュヒットミス判定を行う（ステップ4483）。当該データがキャッシュメモリに保持されていないキャッシュミスの場合には、ディスクアクセス制御部23に対して当該データベースブロックのドライブからキャッシュメモリ22への転送依頼を行う（ステップ4484）。

【0079】

ステップ4483で当該データベースブロックがキャッシュヒットするか、ステップ4484で転送処理が終了すると、キャッシュメモリ22中の当該データベースブロックに、ログレコードに含まれるデータベース更新履歴情報を反映する（ステップ4485）。

ステップ4481からステップ4485までの処理を、当該トランザクションの全ログレコードの処理が完了するまで行う（ステップ4486）。

【0080】

一方、ステップ4481の判定でトランザクションCOMMITログが出現した場合には、当該トランザクションのログについて全て処理を反映し終わっているのでステップ4487へ進み、トランザクションログ管理テーブル268及びトランザクション抽出ログバッファ264から当該トランザクションに関する全ての情報を削除する。

【0081】

以上説明した様に本実施形態のデータベース処理システムによれば、ログ情報

の内、COMMITされたトランザクションのログ情報を用いて更新を行うので、ホストコンピュータのバッファ上で行われたデータベース処理の内容をログ情報によって記憶装置サブシステム上のデータベース領域へ反映させる際の入出力処理を省略することが可能である。

【0082】

(実施形態3)

以下にデータベース領域のデータの更新を、そのデータベース領域のデータに対応する物理デバイス毎に並列に行う実施形態3のデータベース処理システムについて説明する。

【0083】

図11は本実施形態のデータベース処理システムの概略構成を示す図である。本実施形態の処理は、実施形態1及び実施形態2に共通して実施することができる。すなわち、実施形態1の図6におけるログ追跡処理及び実施形態2の図10におけるログ追跡処理において、データベースブロックのデータベース領域ID、ファイルID及びページ番号から、DB-ディスクブロック変換テーブル28を使用して物理ドライブのドライブ番号を取得した後、ドライブ毎に異なるプロセッサに処理を分割して実行することで、キャッシュメモリ22への書き込み処理を並列化する。つまり、キャッシュメモリに書き込むことで、各ドライブへの反映処理を並列処理させることができ、高速なデータ反映が可能となる。ここで本実施形態のディスクサブシステムは、ドライブ毎の処理を実行する複数のプロセッサを備えているものとする。

【0084】

以上説明した様に本実施形態のデータベース処理システムによれば、ログ情報によるデータベース領域のデータの更新を、そのデータベース領域のデータに対応する物理デバイス毎に並列に行うので、ホストコンピュータのバッファ上で行われたデータベース処理の内容をログ情報によって記憶装置サブシステム上のデータベース領域へ反映させる際に、その処理を効率的に行うことが可能である。

【0085】

(実施形態4)

以下にディスクサブシステムを追加した場合の実施形態 4 のデータベース処理システムについて説明する。

図 12 は本実施形態のデータベース処理システムの概略構成を示す図である。図 12 に示す様に本実施形態のディスクサブシステム 1200 はデータ送信処理部 31 を有している。データ送信処理部 31 は、ログ情報の内容に従って更新されるデータベース領域のデータがディスクサブシステム 1200 とは異なるディスクサブシステム 1201 中に存在している場合に、そのログ情報の内容をディスクサブシステム 1201 へ送信する処理部である。

【0086】

ディスクサブシステム 1200 をデータ送信処理部 31 として機能させる為のプログラムは、フロッピーディスク等の記録媒体に記録されて実行されるものとする。なお前記プログラムを記録する記録媒体はフロッピーディスク以外の他の記録媒体でも良い。また前記プログラムを当該記録媒体から情報処理装置にインストールして使用しても良いし、ネットワークを通じて当該記録媒体にアクセスして前記プログラムを使用するものとしても良い。

【0087】

またディスクサブシステム 1201 はデータ受信処理部 32 を有している。データ受信処理部 32 は、前記送信されたログ情報を受信して、ディスクサブシステム 1201 の更新処理部 30 へ、ディスクサブシステム 1201 内のデータベース領域 24 のデータの更新を指示する処理部である。

【0088】

ディスクサブシステム 1201 をデータ受信処理部 32 として機能させる為のプログラムは、フロッピーディスク等の記録媒体に記録されて実行されるものとする。なお前記プログラムを記録する記録媒体はフロッピーディスク以外の他の記録媒体でも良い。また前記プログラムを当該記録媒体から情報処理装置にインストールして使用しても良いし、ネットワークを通じて当該記録媒体にアクセスして前記プログラムを使用するものとしても良い。

【0089】

次に本実施形態について図 12 を用いて説明する。本実施形態では、実施形態

1で示した構成にもう1つのディスクサブシステムを追加した場合の例を表している。つまり、追加したディスクサブシステム1201にはデータベース領域だけが存在する例を示している。

【0090】

本実施形態では、データベース管理処理部10が管理するデータベース領域24がディスクサブシステム1200及び1201に存在している。但し、ログ領域はディスクサブシステム1200にしか存在していない。このような場合、ディスクサブシステム1200のディスク制御処理部21には、ホストコンピュータ1からWRITE要求のあったログブロックを判定し、ディスクサブシステム1201にログブロックを送信するデータ送信処理部31が必要になる。また、ディスクサブシステム1201にはデータ送信処理部31から送信されたデータを受信するデータ受信処理部32が必要となる。

【0091】

複数台のディスクサブシステムを用いて大規模なデータベースシステムを構築した場合、このようなデータベース領域が配置される場合があるが、本実施形態では、以下の様に処理を行う。

すなわち、ディスクサブシステム1200のデータ送信処理部31は、ログ情報の内容に従って更新されるデータベース領域のデータについて、そのディスク制御装置番号を図2のDB-ディスクブロック変換テーブル28から求め、その求めたディスク制御装置番号がディスクサブシステム1200の番号と異なっている場合には、そのディスク制御装置番号のディスクサブシステム、例えばディスクサブシステム1201のデータ受信処理部32へ送信し、データ受信処理部32により、そのログ情報を受信して、ディスクサブシステム1201の更新処理部30へ、ディスクサブシステム1201内のデータベース領域24のデータの更新を指示する。

【0092】

以上説明した様に本実施形態のデータベース処理システムによれば、ログ情報の内容に従って更新されるデータベース領域のデータが他の記憶装置サブシステム中に存在している場合に、そのログ情報の内容を当該記憶装置サブシステムへ

送信し、その記憶装置サブシステム内でデータベース領域のデータの更新を行うので、複数台の記憶装置サブシステムにデータベース領域を配置している大規模なデータベース処理システムで、ホストコンピュータのバッファ上で行われたデータベース処理の内容を記憶装置サブシステム上のデータベース領域へ反映させる際にも、ホストコンピュータと記憶装置サブシステムとの間の入出力処理負荷を低減させることが可能である。また、反映したデータを利用する場合も複数の記憶装置サブシステム内のキャッシュメモリに格納されているため、並列に参照することができ、高速にアクセスすることが可能である。

【0 0 9 3】

【発明の効果】

本発明によればホストコンピュータで行われたデータベース処理の内容を示すログ情報に従って記憶装置サブシステム上のデータベース領域のデータを更新するので、ホストコンピュータで行われたデータベース処理の内容を記憶装置サブシステム上のデータベース領域へ反映させる際に、ホストコンピュータと記憶装置サブシステムとの間の入出力処理負荷を低減させることが可能である。

【図面の簡単な説明】

【図 1】

実施形態 1 のデータベース処理システムのシステム構成を示す図である。

【図 2】

実施形態 1 の D B - ディスクブロック変換テーブル 2 8 の構成情報を示す図である。

【図 3】

実施形態 1 のホストコンピュータ上で認識されるデータベース領域と、オペレーティングシステムが認識する論理ボリュームと、デバイスファイル及びディスクサブシステム内の L U へのマッピング関連の例を示す図である。

【図 4】

実施形態 1 の受信コマンド解析処理の処理手順を示すフローチャートである。

【図 5】

実施形態 1 の W R I T E コマンド受領時の処理手順を示すフローチャートであ

る。

【図 6】

実施形態 1 のログ追跡処理の処理手順を示すフローチャートである。

【図 7】

実施形態 1 の READ コマンド受領時の処理手順を示すフローチャートである。

【図 8】

実施形態 2 のログ追跡処理においてログブロック中に含まれる全てのログレコードについてトランザクション毎にログを解析した結果の例を示す図である。

【図 9】

実施形態 2 の図 8 のトランザクションログ管理テーブル 268 を用いる際のログレコード分別処理の処理手順を示すフローチャートである。

【図 10】

実施形態 2 の図 9 のステップ 448 におけるログ追跡処理の処理手順を示すフローチャートである。

【図 11】

実施形態 3 のデータベース処理システムの概略構成を示す図である。

【図 12】

実施形態 4 のデータベース処理システムの概略構成を示す図である。

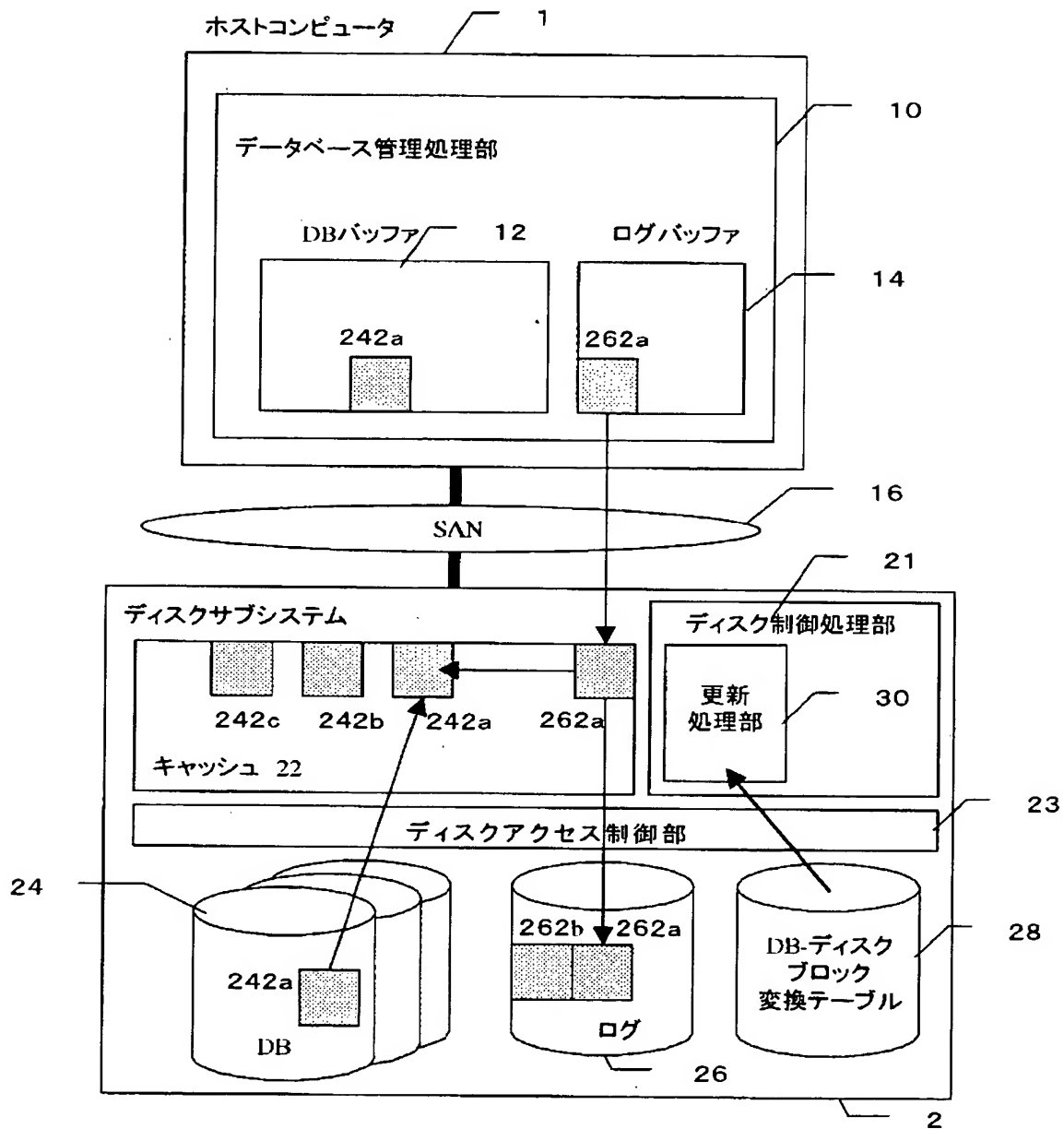
【符号の説明】

1…ホストコンピュータ、2…ディスクサブシステム、12…DBバッファ、14…ログバッファ、16…ストレージエリアネットワーク、22…キャッシュメモリ、24…データベース領域、242…データベースブロック、26…ログ用ディスク、262…ログブロック、28…DB-ディスクブロック変換テーブル、10…データベース管理処理部、21…ディスク制御処理部、23…ディスクアクセス制御部、30…更新処理部、264…抽出ログバッファ、266…ログレコード、268…トランザクションログ管理テーブル、1200及び1201…ディスクサブシステム、31…データ送信処理部、32…データ受信処理部。

【書類名】 図面

【図 1】

図 1



【図 2】

図 2

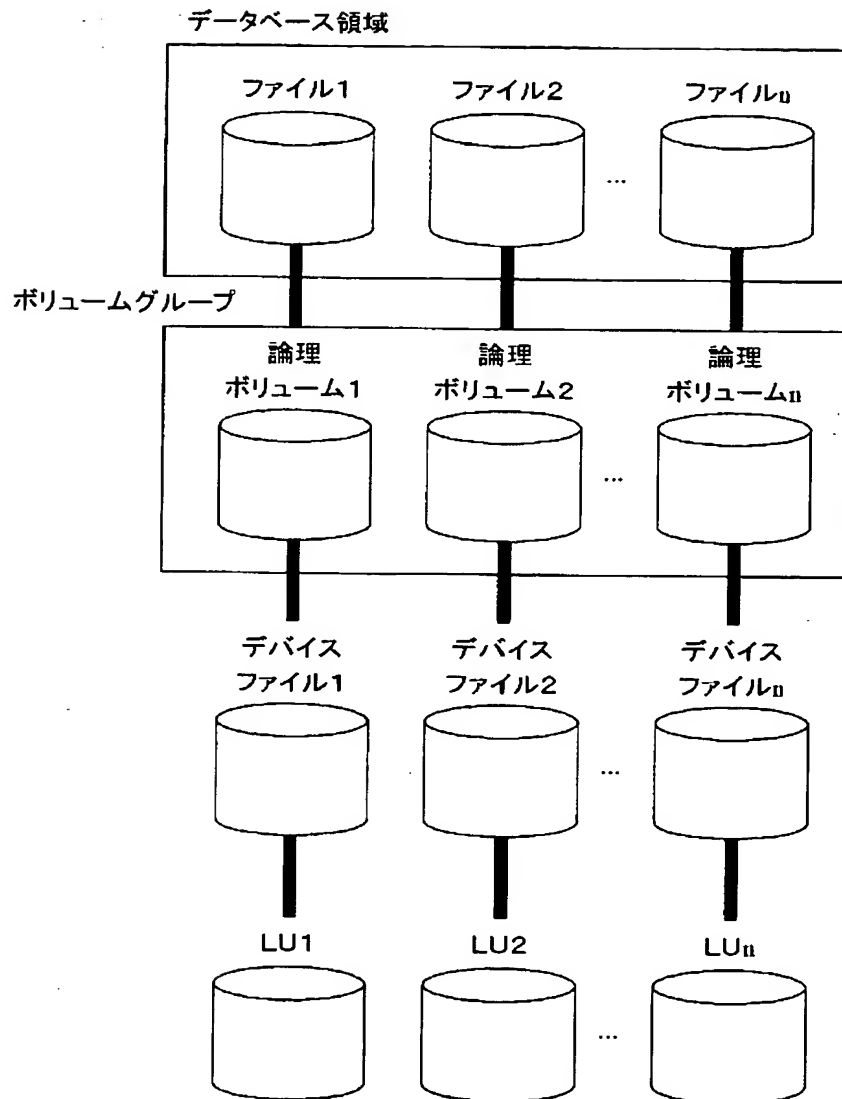
DB-ディスクブロック変換テーブル

データベース領域ID	ファイルID	ブロック長	論理ボリュームID	ディスク 制御装置番号	物理デバイスID (LUN)	相対位置 (LBA)

28

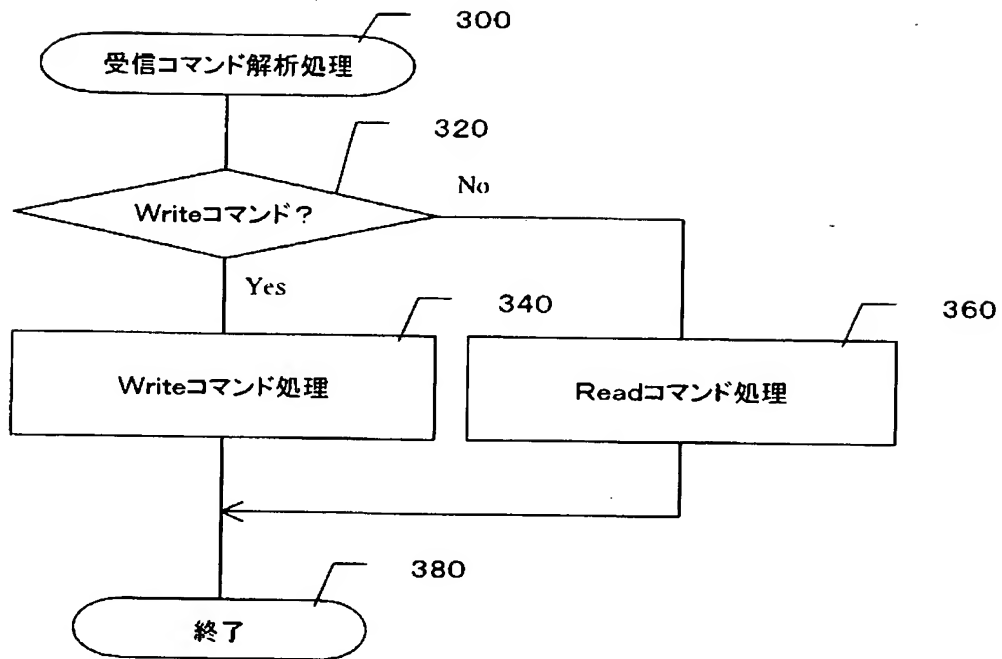
【図 3】

図 3



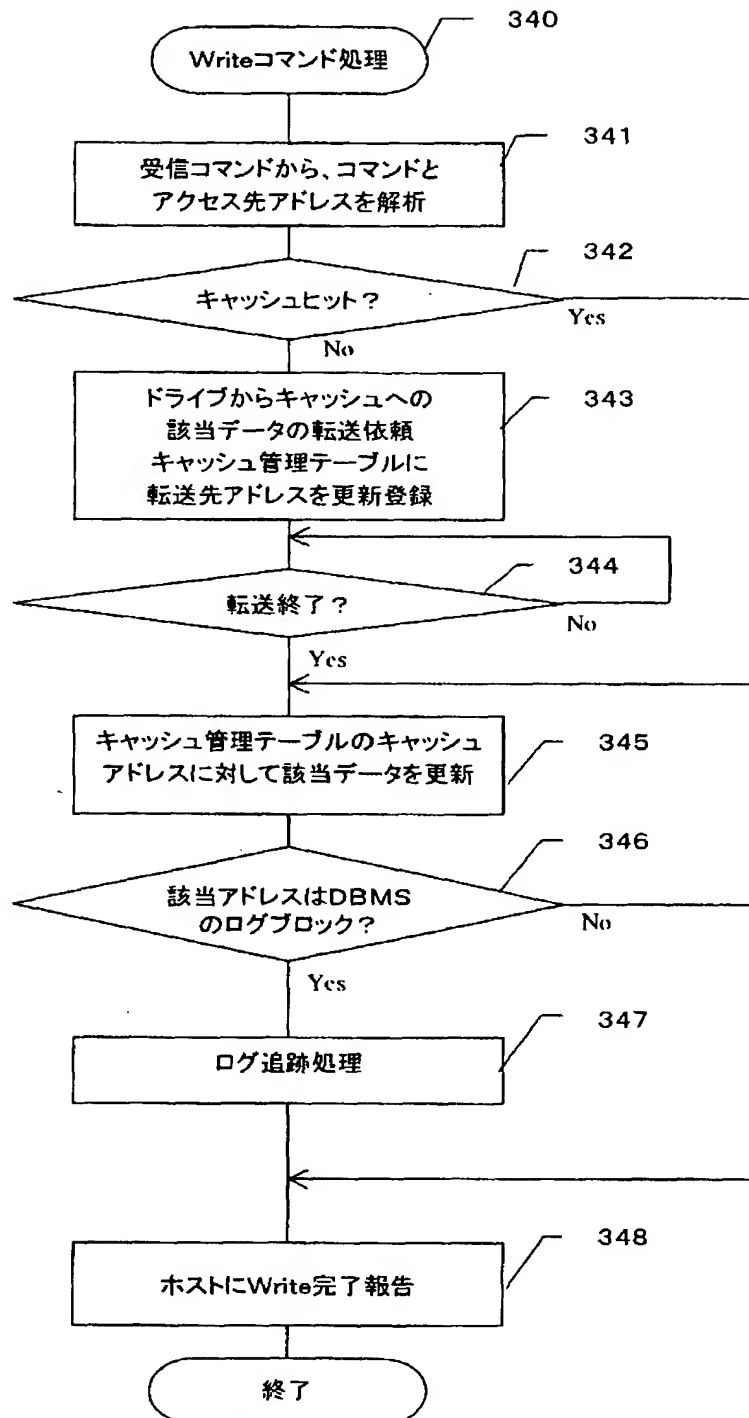
【図 4】

図 4

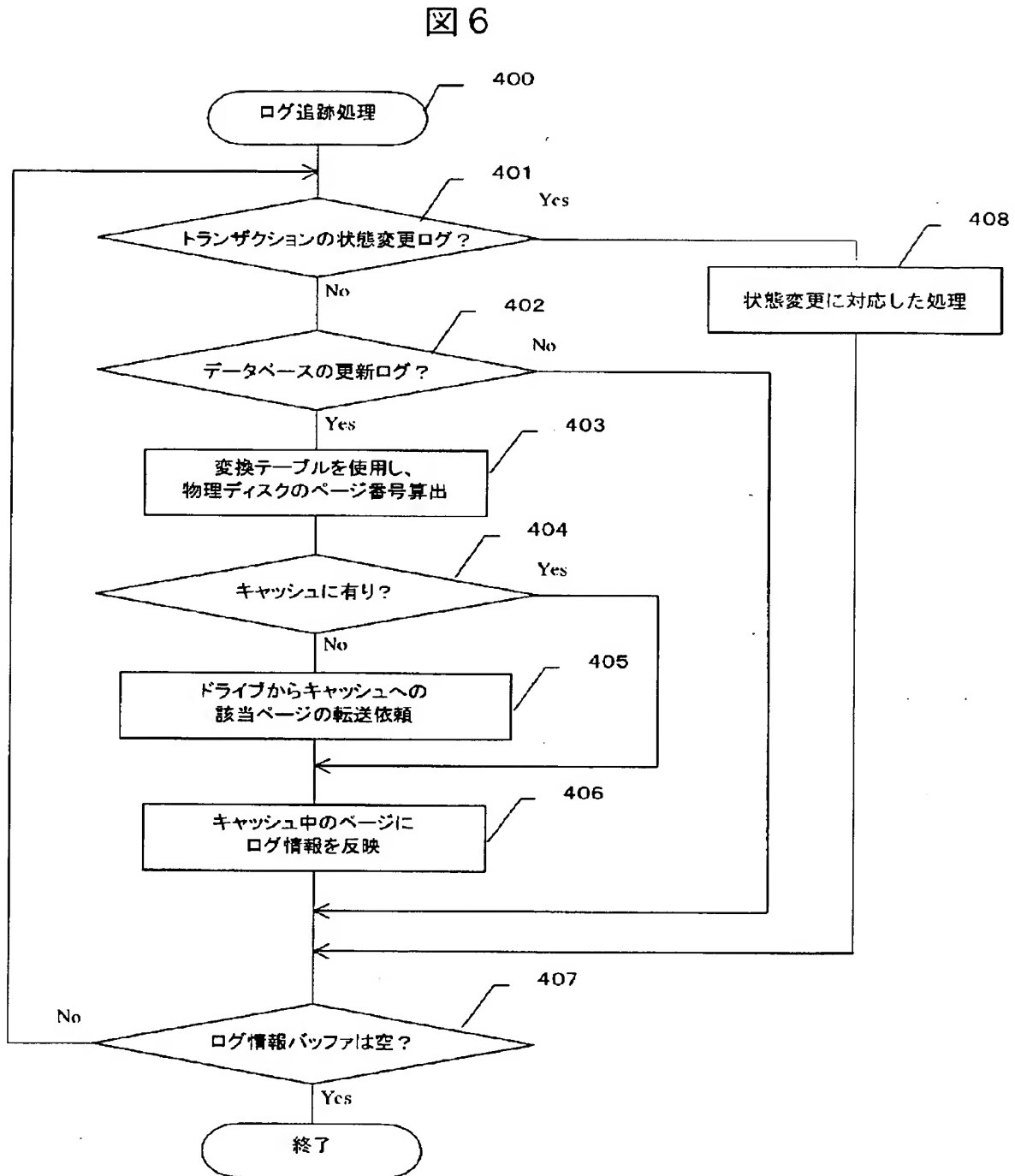


【図 5】

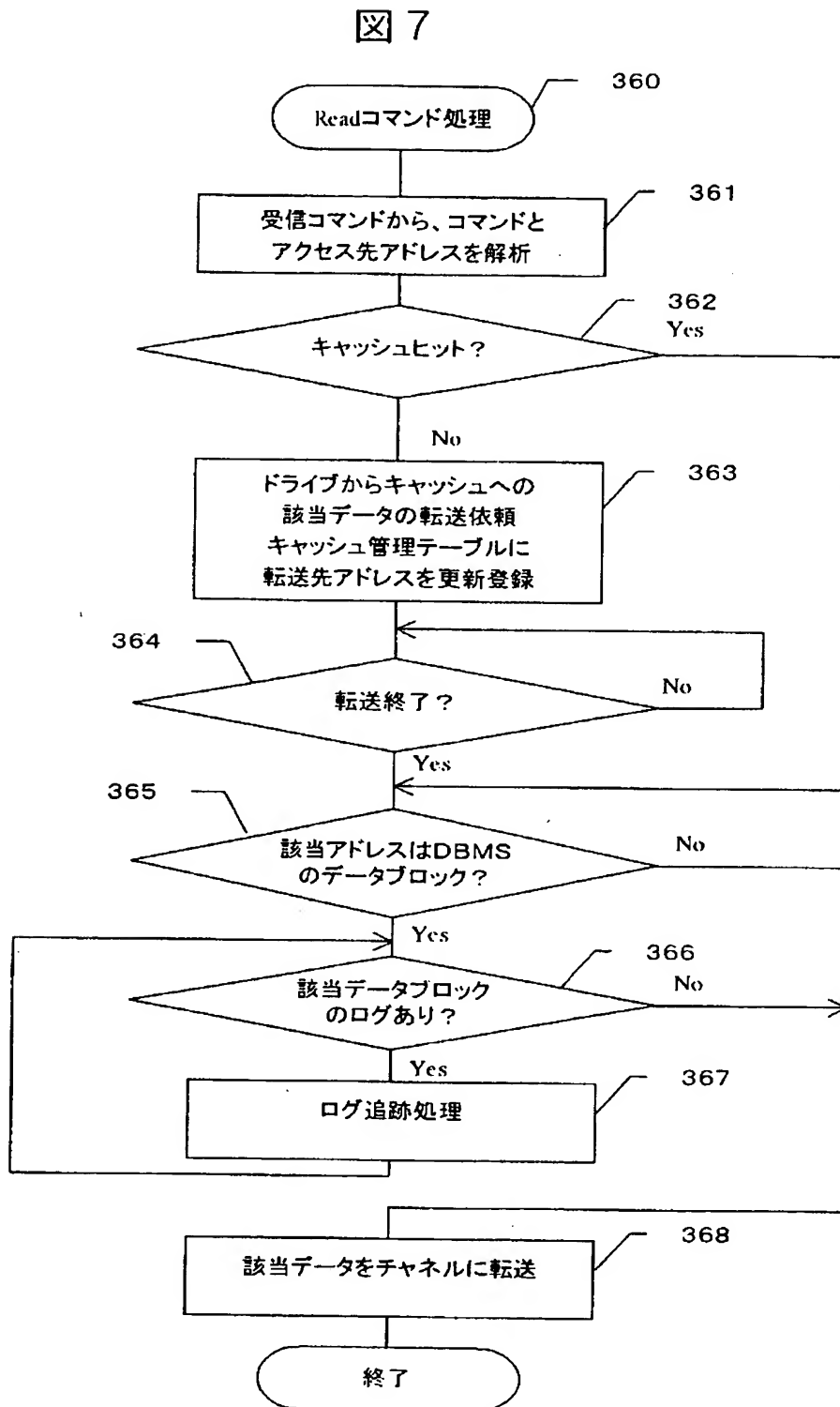
図 5



【図 6】

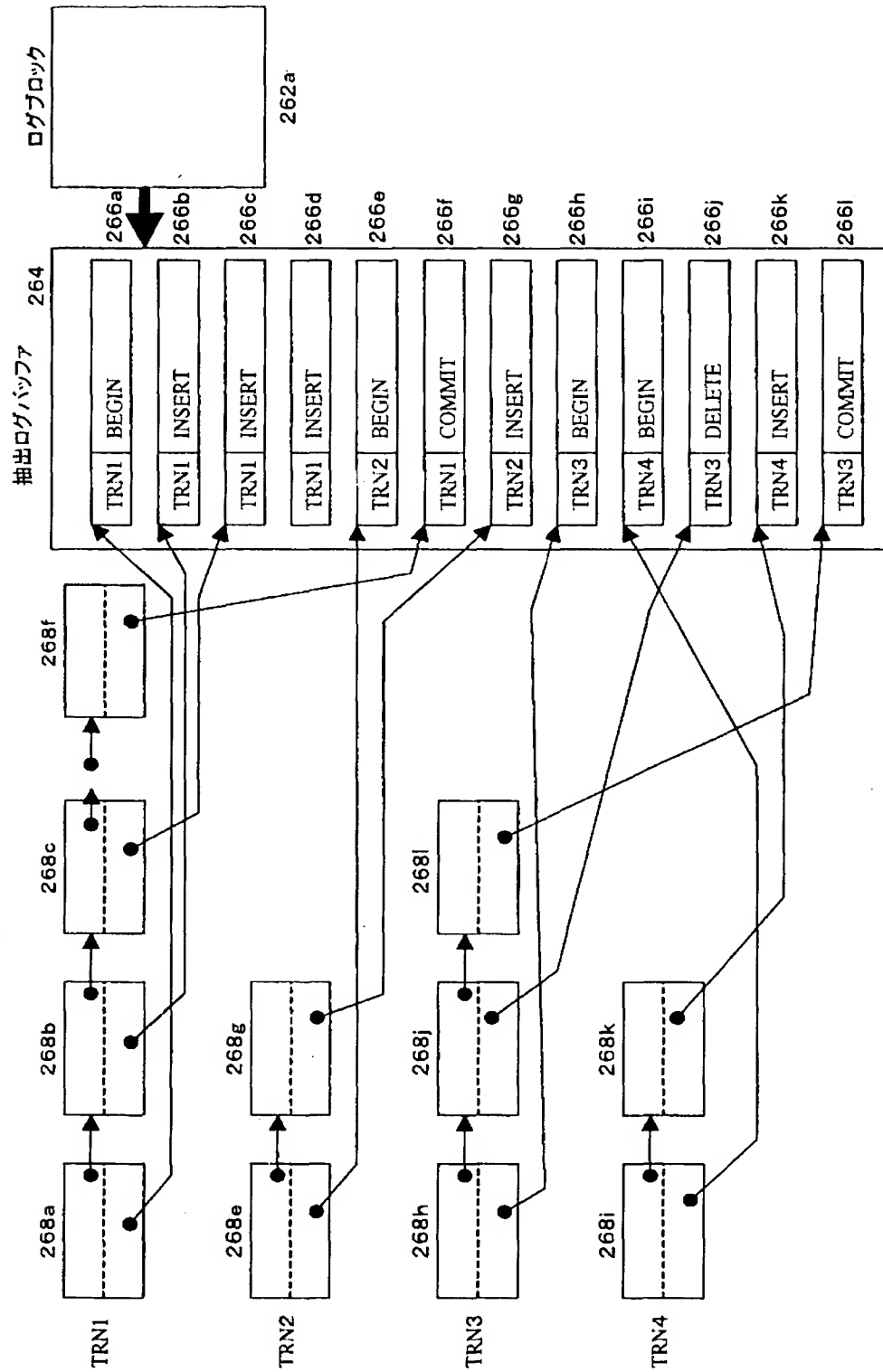


【図 7】



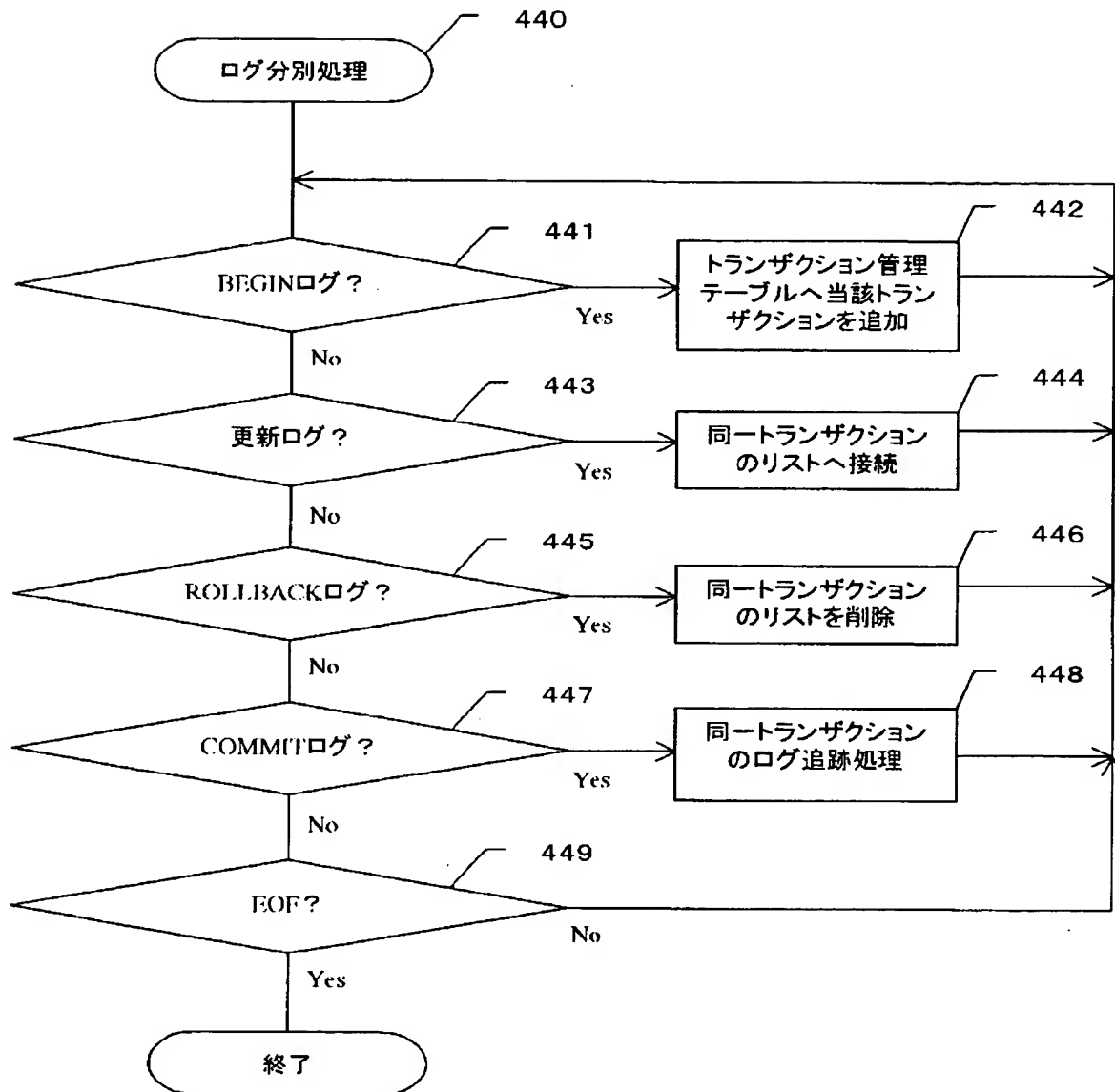
【図 8】

図 8
トランザクションログ管理テーブル

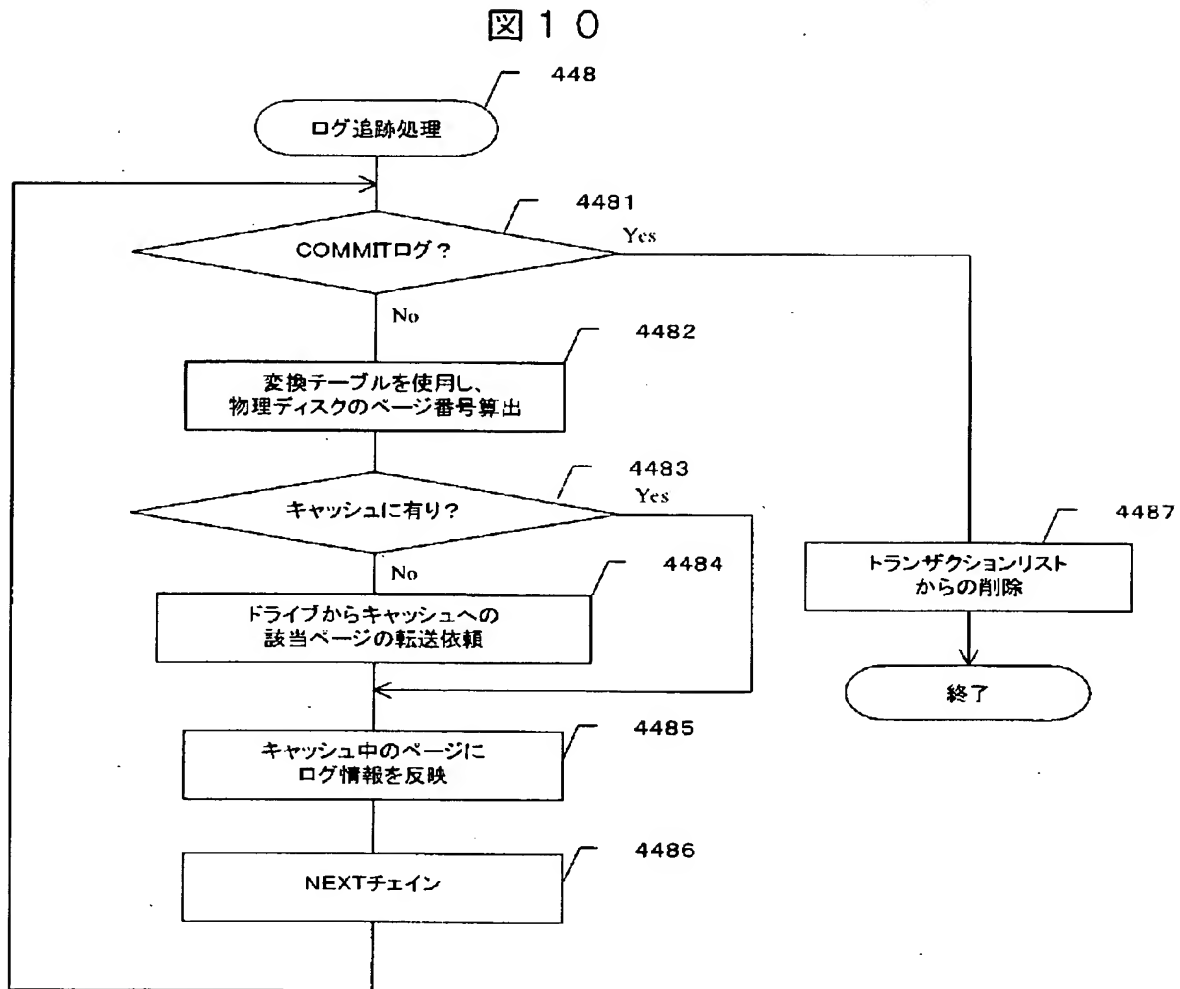


【図 9】

図 9

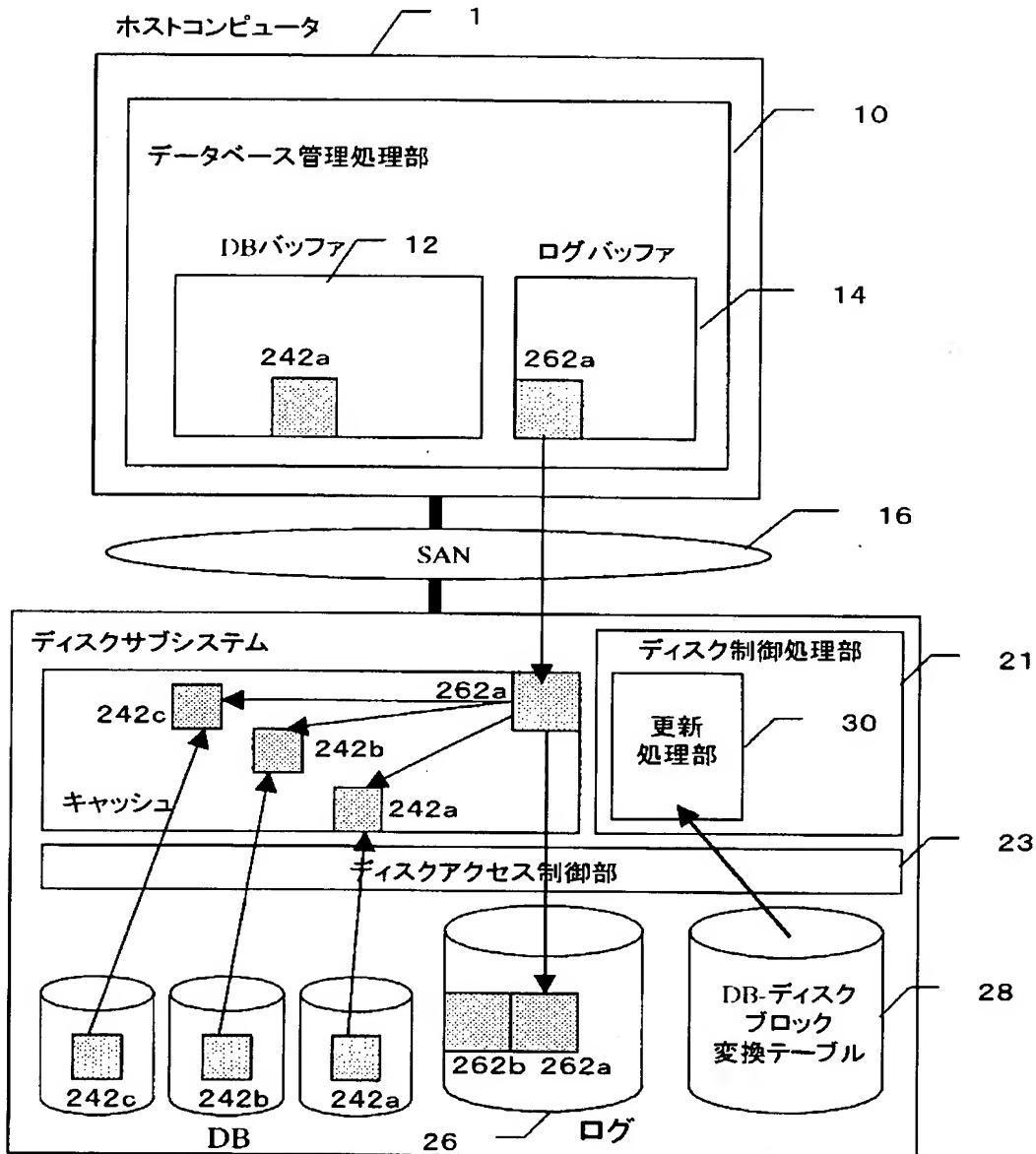


【図 10】

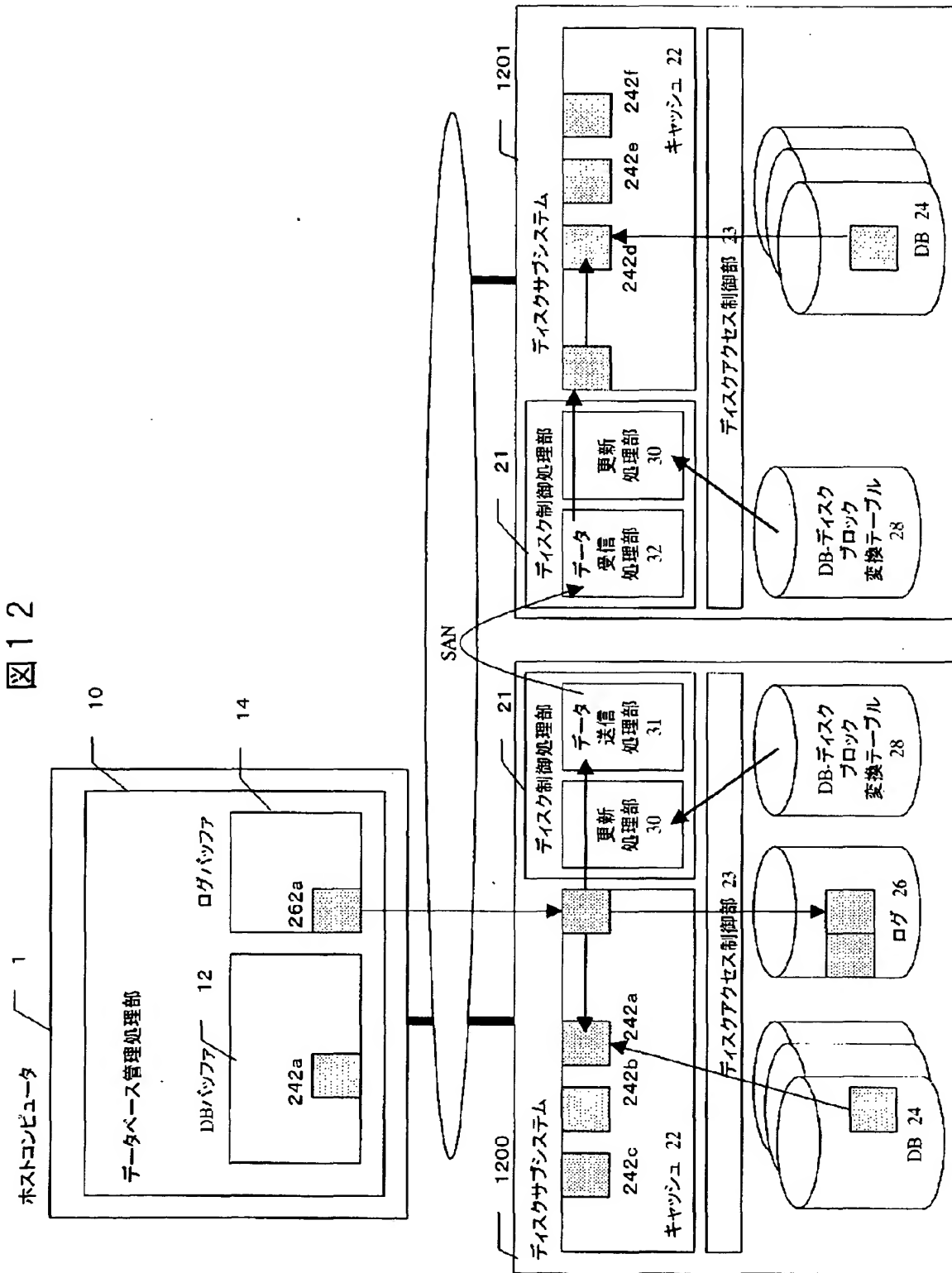


【図 11】

図 11



【図 12】



【書類名】 要約書

【要約】

【課題】 ホストコンピュータのバッファの内容を記憶装置サブシステム上のデータベース領域へ反映させる際に入出力処理負荷を低減させる。

【解決手段】 ホストコンピュータからのアクセス要求が書き込み要求であり、その書き込み内容がホストコンピュータのバッファ上で行われたデータベース処理の内容を示すログ情報であるかどうかを判定するステップと、前記書き込み内容が前記ログ情報である場合に、ホストコンピュータ側のデータベース処理で認識している論理的な位置情報と記憶装置サブシステム上の物理的な位置情報との対応関係を示す変換テーブルによって、前記ログ情報中に示された位置情報を記憶装置サブシステム上の物理的な位置情報に変換するステップと、その変換した物理的な位置情報で表される記憶装置サブシステム上のデータベース領域のデータを前記ログ情報の内容に従って更新するステップとを有するものである。

【選択図】 図 1

認定・付加情報

特許出願の番号	特願 2 0 0 2 - 3 6 8 6 8 8
受付番号	5 0 2 0 1 9 2 9 6 7 5
書類名	特許願
担当官	第七担当上席 0 0 9 6
作成日	平成 1 5 年 1 月 6 日

< 認定情報・付加情報 >

【提出日】	平成 14 年 12 月 19 日
-------	-------------------

次頁無

特願 2 0 0 2 - 3 6 8 6 8 8

出 願 人 履 歴 情 報

識別番号

[0 0 0 0 5 1 0 8]

1. 変更年月日

1 9 9 0 年 8 月 3 1 日

[変更理由]

新規登録

住 所

東京都千代田区神田駿河台 4 丁目 6 番地

氏 名

株式会社日立製作所